



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

Too sophisticated even for highly educated survey respondents? A qualitative assessment of indirect question formats for sensitive questions

Jerke, Julia ; Johann, David ; Rauhut, Heiko ; Thomas, Kathrin

Abstract: Misreporting of sensitive characteristics in surveys is a major concern among survey methodologists and social scientists across disciplines. Indirect question formats, such as the Item Count Technique (ICT) and the Randomized Response Techniques (RRT), including the Crosswise Model (CM) and the Triangular Model (TM), have been developed to protect respondents' privacy by design to elicit more truthful answers. These methods have also been praised to produce more valid estimates than direct questions. However, recent research has revealed a number of problems, such as the occurrence of false negatives, false positives, and dependencies on socioeconomic characteristics, indicating that at least some respondents may still cheat or lie when asked indirectly. This article systematically investigates (1) how well respondents comprehend and (2) to what extent they trust the ICT, CM and TM. We conducted cognitive interviews with academics across disciplines, investigating how respondents perceive, think about and answer questions on academic misconduct using these indirect methods. The results indicate that most respondents comprehend the basic instructions, but many fail to understand the logic and principles of these techniques. Furthermore, the findings suggest that comprehension and honest self-reports are unrelated, thus violating core assumptions about the effectiveness of these techniques.

DOI: <https://doi.org/10.18148/srm/2019.v13i3.7453>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-180713>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Jerke, Julia; Johann, David; Rauhut, Heiko; Thomas, Kathrin (2019). Too sophisticated even for highly educated survey respondents? A qualitative assessment of indirect question formats for sensitive questions. *Survey Research Methods*, 13(3):319-351.

DOI: <https://doi.org/10.18148/srm/2019.v13i3.7453>

Too sophisticated even for highly educated survey respondents? A qualitative assessment of indirect question formats for sensitive questions

Julia Jerke
University of Zürich

David Johann
University of Zürich

Heiko Rauhut
University of Zürich

Kathrin Thomas
Princeton University

Misreporting of sensitive characteristics in surveys is a major concern among survey methodologists and social scientists across disciplines. Indirect question formats, such as the Item Count Technique (ICT) and Randomized Response Techniques (RRT), including the Crosswise Model (CM) and the Triangular Model (TM), have been developed to protect respondents' privacy by design to elicit more truthful answers. These methods have also been praised to produce more valid estimates than direct questions. However, recent research has revealed a number of problems, such as the occurrence of false negatives, false positives, and dependencies on socioeconomic characteristics, indicating that at least some respondents may still cheat or lie when asked indirectly. This article systematically investigates (1) how well respondents comprehend and (2) to what extent they trust the ICT, CM and TM. We conducted cognitive interviews with academics across disciplines, investigating how respondents perceive, think about and answer questions on academic misconduct using these indirect methods. The results indicate that most respondents comprehend the basic instructions, but many fail to understand the logic and principles of these techniques. Furthermore, the findings suggest that comprehension and honest self-reports are unrelated, thus violating core assumptions about the effectiveness of these techniques.

Keywords: Item Count Technique; Crosswise Model; Triangular Model; Cognitive Interviews; Academic Misconduct

1 Introduction

Many surveys capture sensitive characteristics in addition to less sensitive attitudinal or factual questions. Thus, an increasing body of literature is dedicated to researching how to improve questions capturing sensitive characteristics, as these items are prone to misreporting (e.g., Krumpal, Jann, Auspurg, & von Hermann, 2015; Tourangeau, Rips, & Rasinski, 2000; Tourangeau & Yan, 2007). In particular, indirect question formats aim to reduce social desirability pressures by protecting respondents' anonymity and privacy. This is achieved by design implementing a statistical element that allows an analysis of the collected data on the aggregate

level using basic probability methods.

Three popular methods are the Item Count Technique (ICT) (Droitcour et al., 1991),¹ the Crosswise Model and the Triangular Model (CM and TM) (Yu, Tian, & Tang, 2008). The common elements of the three methods are that they (1) obscure the meaning of respondents' individual answers; (2) they do this by design and without additional aids; and (3) the resulting data only allows an estimate of the prevalence of sensitive characteristics on the aggregated level.

Although these methods are frequently applied to measure sensitive characteristics, an increasing body of literature raises concerns about their effectiveness (Coutts, Jann, Krumpal, & Näher, 2011; Droitcour et al., 1991; Hoffmann, de Puiseau, Schmidt, & Musch, 2017; Höglinger & Diekmann, 2017; Höglinger & Jann, 2018; Landsheer, Van Der Heijden, & Van Gils, 1999; Wolter, 2012; Wolter & Laier, 2014). Many of the studies exploring the efficacy of indirect questioning techniques seem to base their investigations on three assumptions: (1) respondents comprehend the meth-

Contact information: Julia Jerke, Institute of Sociology, Andreasstrasse 15, 8050 Zürich, Switzerland (E-Mail: jerke@soziologie.uzh.ch)

Editor's note: This paper deals with small n and highly sensitive information. Since the potential risk of disclosure is considered too high, SRM grant an exception from the requirement to publish replication materials.

¹In the literature the ICT is also commonly referred to as list experiment or unmatched count technique.

ods; (2) correctly follow the instructions; and (3) answer truthfully. However, recent research raises doubt about these assumptions, suggesting the occurrence of false negatives or false positives, which may be the result of the design in general and/or a consequence of the respondents' cognitive processing of the question formats (Hoffmann et al., 2017; Höglinger & Diekmann, 2017; Höglinger & Jann, 2018; Kirchner, 2015; Moshagen, Hilbig, Erdfelder, & Moritz, 2014; Van Der Heijden, Gils, Bouts, & Hox, 2000; Wolter, 2012).

Arguably, the ICT, CM and TM impose a higher cognitive burden on respondents than direct questions. While their design is more complex, given long instructions and dense information, respondents have little time to respond in the survey situation (e.g., Fowler Jr, 1995; Fowler Jr & Cosenza, 2009). Previous research has studied whether respondents understand and trust Randomized Response Techniques (RRT) and ICT formats (e.g., Coutts et al., 2011; Höglinger & Diekmann, 2017; Höglinger & Jann, 2018; Wolter, 2012; Wolter & Laier, 2014). Only few studies have investigated the direct link between comprehension, trust and honesty, as well as the validity of these indirect questioning techniques (Droitcour et al., 1991; Hoffmann et al., 2017; Landsheer et al., 1999; Wolter, 2012). The evidence is mixed: For example, while Landsheer et al. (1999) demonstrate that a better understanding of RRT corresponds with socially undesirable responses, Wolter (2012) finds no significant connection between comprehension and trust in protection of anonymity as well as the number of valid answers. Moreover, Hoffmann et al. (2017) suggest that RRT questions are generally well comprehended, but find little correlation between comprehension and the perceived level of privacy protection. Finally, cognitive interviews conducted by Droitcour et al. (1991) reveal that respondents mostly understand how to answer questions based on the ICT, but that not all of them understand the purpose of the method.

Our article directly ties in with this work. We investigate how well respondents comprehend the instructions of the ICT, CM and TM and the purpose of asking questions in this way. Moreover, we examine whether a higher level of comprehension is related to a higher level of trust in these methods. We test this by applying a research design that is qualitative in nature: We study respondents' thinking processes when answering questions on academic misconduct, using ICT, CM and TM. We employ cognitive interviews with 19 academics in three countries (Germany, the United Kingdom and Switzerland), varying in their academic discipline, academic status, and gender.

Focusing on academic misconduct has a number of advantages: Firstly, it is a reasonable assumption that highly educated respondents, such as academics, should be more qualified to understand sophisticated question formats and the logic behind them. If academics find it hard to understand

the mechanisms of the ICT, CM and TM, it is highly likely that those with lower education struggle, too (e.g., Hoffmann et al., 2017). As such, we believe that academic respondents set a good benchmark. Secondly, misconduct is a major concern in academia, especially considering increasing pressure to "publish or perish" (e.g., Franco, Malhotra, & Simonovits, 2014). Solutions to capture academic misconduct better are thus important for future research in this field. Finally, especially studies using the CM and TM have predominantly been applied to student samples (Schnell, Thomas, & Noack, 2019).

We proceed as follows: We first discuss the ICT, CM and TM designs in more detail. Next, we review research on the cognitive mechanisms of the survey response to determine how the three methods may be affected by respondents' thinking processes. We then discuss our study design before presenting the results of our research. We close with a discussion of the implications of this research for survey methodology and social science in general, and the field of academic misconduct in particular.

2 ICT, CM and TM

To recap, indirect question formats, especially RRT, were developed to address issues of misreporting in the survey situation. While many different variants of those designs have been proposed (Boruch, 1971; Clark & Desharnais, 1998; Droitcour et al., 1991; Greenberg, Abul-Ela, Simmons, & Horvitz, 1969; Horvitz, Simmons, & Shah, 1968; Krumpal et al., 2015; Kuk, 1990; Lensvelt-Mulders, Hox, Van der Heijden, & Maas, 2005; Umesh & Peterson, 1991; Warner, 1965), we focus on three popular methods in the recent literature: the ICT (e.g., Droitcour et al., 1991), CM and TM (e.g., Yu et al., 2008), which we discuss in more detail in the following.

2.1 On the logic of the ICT

The simplest variant of the ICT (Droitcour et al., 1991) is based on a split ballot design, in which one subsample is presented with a short list of unobtrusive behaviours and the other with the same list of unobtrusive behaviours plus the sensitive characteristic of interest (Holbrook & Krosnick, 2010). Respondents in both subsamples are then requested to state how many of the statements apply to them, as opposed to reporting individual answers to each list item.

The prevalence π of the sensitive characteristic is then given by the mean difference of the long list and the short list

$$\hat{\pi}_{ICT} = \bar{X}_{\text{long list}} - \bar{X}_{\text{short list}} \quad (1)$$

and the sampling variance $\text{Var}(\hat{\pi}_{ICT})$ is calculated as follows (Tourangeau & Yan, 2007):

$$\text{Var}(\hat{\pi}_{ICT}) = \text{Var}(\bar{X}_{\text{long list}}) + \text{Var}(\bar{X}_{\text{short list}}). \quad (2)$$

Alternative designs have been developed (Droitcour et al., 1991; Glynn, 2013; Grant, Moon, & Gleason, 2012, 2014), but the traditional ICT remains the most commonly used design. While its simple implementation is one core advantage, it appears to be more challenging to create an appropriate experimental design. For example, scholars may underestimate the importance of the selection of the unobtrusive items, which need to cover up the sensitive characteristic sufficiently (Tsuchiya, Hirai, & Ono, 2007): To avoid the loss of privacy protection due to floor and ceiling effects, which arise when respondents either disagree or agree with all items on the list, researchers need to ensure enough variation in the prevalence of the unobtrusive items.² It is recommended to include at least one item with a very low prevalence and one item with a very high prevalence, along with a few items that vary in the frequency with which they occur.

Among other things, prior research has implemented the ICT to study theft, shoplifting, crime, sexual preferences, attitudes to minority groups, and voting behaviour (e.g., Biemer & Brown, 2005; Comşa & Postelnicu, 2012; Dalton, Wimbush, & Daily, 1994; Díaz-Cayeros, Magaloni, Matanock, & Romero, 2011; Gilens, Sniderman, & Kuklinski, 1998; Holbrook & Krosnick, 2010; LaBrie & Earleywine, 2000; Rayburn, Earleywine, & Davison, 2003; Thomas, Johann, Kritzing, Plescia, & Zeglovits, 2016; Tsuchiya et al., 2007; Wolter & Laier, 2014). The results suggest that the ICT might at least be an appropriate method to capture socially undesirable, infrequently practised behaviour. However, it seems to be less effective in capturing socially desirable behaviour, eventually understating social desirability bias when high-incidence desirable behaviour is captured (De Jonge & Nickerson, 2014; Thomas et al., 2016).

2.2 The logic of the CM and TM

Many RRT rely on randomization devices to obscure the survey responses in a way so that neither interviewers nor researchers are able to identify whether or not the sensitive characteristic applies to the particular respondent (Boruch, 1971; Clark & Desharnais, 1998; Greenberg et al., 1969; Horvitz et al., 1968; Krumpal et al., 2015; Kuk, 1990; Lensvelt-Mulders et al., 2005; Umesh & Peterson, 1991; Warner, 1965). This is possible because the probabilities of the randomization device are known and simple probability methods allow one to estimate the prevalence of the sensitive characteristic. However, the use of auxiliary devices, such as coins or dice, in survey situations may result in refusals, break-offs or in dishonest or self-protective ‘no’-answers (Krumpal et al., 2015; Ulrich, Schröter, Striegel, & Simon, 2012). The CM and TM overcome the problem of using auxiliary devices in surveys by implementing randomization in the question design (Yu et al., 2008).

Both techniques rely on two simple questions, a non-

sensitive one with known probabilities, typically asking whether a close person’s birthday (e.g., mother, father or friend) is in a particular month (e.g., October, November or December),³ and a question about the sensitive characteristic. Respondents are instructed to provide only one joint answer to both questions (Höglinger, Jann, & Diekmann, 2016; Korndörfer, Krumpal, & Schmukle, 2014; Ulrich et al., 2012).

For the CM respondents choose between code (A), meaning the answers to the two questions are both ‘no’ or both ‘yes’, and code (B), meaning the answers to the two questions differ, i.e., one is ‘no’ and one is ‘yes’. The TM potentially offers more protection allowing a code (A), meaning the answer to both questions is ‘no’, and a code (B), meaning the answer is ‘yes’ to at least one of the questions. Given that (1) the response codes to both questions are binary; (2) the non-sensitive behaviour is unrelated to the sensitive one; and (3) the non-sensitive behaviour has a known probability p , it is possible to estimate the prevalence of the sensitive characteristic on the sample (Krumpal et al., 2015).

The prevalence π of the sensitive behaviour in the CM condition can be estimated by

$$\hat{\pi}_{CM} = \frac{\hat{\lambda} + p - 1}{2p - 1}, p \neq 0.5, \quad (3)$$

where p is the known population prevalence of the non-sensitive item (Yu et al., 2008). Assuming a uniform (Yu et al., 2008) birthday distribution, the probability of being born in October, November or December is $p = 0.25$ (Jann, Jerke, & Krumpal, 2012). $\hat{\lambda}$ is the proportion of respondents stating that their answer to both the birthday and the sensitive question is the same (option (A)). The sampling variance $\text{Var}(\hat{\pi}_{CM})$ is calculated as follows:

$$\begin{aligned} \text{Var}(\hat{\pi}_{CM}) &= \frac{\hat{\lambda}(1 - \hat{\lambda})}{n(2p - 1)^2} \\ &= \frac{\hat{\pi}_{CM}(1 - \hat{\pi}_{CM})}{n} + \frac{p(1 - p)}{n(2p - 1)^2}, p \neq 0.5. \end{aligned} \quad (4)$$

²Floor effects, i.e., reconstructing a list with low prevalence items, are problematic as respondents may be alerted by a the low prevalence list – many may have engaged in none of the behaviours, which causes suspicion. Ceiling effects, i.e., constructing a list with only high prevalence items, may allow that the sensitive characteristic is revealed and encourage respondents to answer dishonestly (Glynn, 2013). Research on the ICT suggests that floor and ceiling effects may occur, which can bias the resulting estimate. While methods have been proposed to statistically control and test for these effects (G. Blair & Imai, 2012), they remain an issue that essentially needs to be addressed and carefully evaluated at the design stage (G. Blair, Imai, & Park, 2010; Glynn, 2013; Tsuchiya et al., 2007).

³Other questions with known probabilities, such as Benford’s law looking at house numbers have been tested (Kundt, 2014; Schnell et al., 2019).

By contrast, the prevalence π of the sensitive behaviour in the TM can be estimated by

$$\hat{\pi}_{\text{TM}} = 1 - \frac{\hat{\lambda}}{1 - p}, \quad (5)$$

where p is the known population prevalence of the non-sensitive item (in the birthday example, $p = 0.25$) and $\hat{\lambda}$ is the proportion of respondents stating that their answer to both the birthday and the sensitive question is ‘no’ (option (A)). The sampling variance $\text{Var}(\hat{\pi}_{\text{TM}})$ is calculated as follows (Yu et al., 2008):

$$\text{Var}(\hat{\pi}_{\text{TM}}) = \frac{\hat{\pi}_{\text{TM}}(1 - \hat{\pi}_{\text{TM}})}{n} + \frac{p(1 - \hat{\pi}_{\text{TM}})}{n(1 - p)}. \quad (6)$$

The results regarding the effectiveness of the CM and TM are mixed. Many previous studies using the CM or the TM posit that the methods significantly reduce under-reporting of sensitive characteristics, such as plagiarism, cheating in games, substance abuse, sexual behaviour, tax evasion, undeclared employment, shoplifting, Xenophobia and Islamophobia (Coutts et al., 2011; Hoffmann, Diedenhofen, Verschuere, & Musch, 2015; Hoffmann & Musch, 2016; Höglinger & Jann, 2018; Höglinger et al., 2016; Jann et al., 2012; Jerke & Krumpal, 2013; Johann & Thomas, 2017; Korndörfer et al., 2014; Krumpal, 2012; Kundt, 2014; Kundt, Misch, & Nerré, Birger, 2017; Nakhaee, Pakravan, & Nakhaee, 2013; Shamsipour et al., 2014; Vakilian, Abbas Mousavi, Keramat, & Chaman, 2016; Vakilian, Mousavi, & Keramat, 2014). However, this does not seem to apply to all tested items and studies. Some research indicates small and statistically insignificant differences for items in the same topical areas (Coutts et al., 2011; Höglinger & Jann, 2018; Jann et al., 2012; Johann & Thomas, 2017; Kundt et al., 2017; Schnell et al., 2019; Shamsipour et al., 2014). Furthermore, Schnell et al. (2019) noted that many studies are implemented on non-probability samples, suggesting that the effectiveness of these methods has yet to be confirmed when probability samples of general populations are concerned.

2.3 On the effectiveness of ICT, CM and TM

Even though a large number of studies have arguably produced more valid estimates than direct questions, doubt has been voiced that the methods may not work as well as previously anticipated and produce false negatives, i.e., respondents incorrectly denying to carry the sensitive trait, or false positives, i.e., respondent incorrectly stating they do carry the sensitive trait. This can be tested in individual-level validation studies that allow a matching of the estimate derived on the basis of indirect questioning with individual respondents’ ‘true’ value. This research demonstrates that indirect questioning techniques are prone to false negatives, which increases with severity of the sensitive trait (Edgell, Duchan,

& Himmelfarb, 1992; Edgell, Himmelfarb, & Duchan, 1982; Höglinger & Jann, 2018; Kirchner, 2015; Lensvelt-Mulders et al., 2005; Moshagen et al., 2014; Van Der Heijden et al., 2000; Wolter, 2012).

Individual-level validation studies are the exception, as they require administrative data that record the ‘true’ status of the respondent. However, prior research also studied whether core assumptions of indirect questioning techniques indeed apply (e.g., Hoffmann et al., 2017; Wolter, 2012). The ICT, CM or TM generally rely on three fundamental assumptions: respondents (1) comprehend the instructions; (2) genuinely follow them; and (3) answer truthfully. When answering indirect questions, respondents are cognitively challenged by an interplay of comprehension, trust and honesty: Respondents who do not fully understand the instructions may find it difficult to trust the methods, while trust in the methods is a necessary condition to answer honestly.

The occurrence of false negatives can be the result of either unknowingly deviating from the rules, commonly referred to as cheating, or deliberately giving an incorrect answer, commonly referred to as lying (Campbell, 1987; Clark & Desharnais, 1998; Lensvelt-Mulders & Boeije, 2007; Ostapczuk, Moshagen, Zhao, & Musch, 2009).⁴

Cheating can be the result of a lack of comprehension. If the instructions for RRT designs are too complicated for the respondents, they may disregard the rules and give arbitrary answers. Studies directly measuring the proportion of respondents correctly understanding the instructions and corresponding rules demonstrate that comprehension rates vary widely and may result in false negatives (Coutts et al., 2011; Hoffmann et al., 2017; Locander, Sudman, & Bradburn, 1976; Miller, 1984). The incentive for cheating should also be higher in indirect questioning, especially if the design does not allow a clear option for respondents to deny carrying the sensitive characteristic (Fox & Tracy, 1980).

Cheating due to a lack of comprehension may also result in the occurrence of false positives, i.e., respondents stating they carry the sensitive trait even though they do not (Höglinger & Diekmann, 2017). False positives are harmful as they raise concerns about the validity of the commonly applied more-is-better assumption (Höglinger & Jann, 2018; Umesh & Peterson, 1991), which suggests that results measured on the basis of indirect questions enquiring about socially undesirable behaviour are more valid, if they yield higher estimates than direct questions. Recent research on the CM suggests a false-positives rate ranging between 5% and 8%, depending on the item attributed to arbitrary answering patterns or cheating (Höglinger & Diekmann, 2017).

Even though respondents may comprehend the instruc-

⁴Our definitions of cheating and lying correspond with Lensvelt-Mulders and Boeije (2007, p. 591) who characterize cheating as “not operating according to randomized response rules” and lying as “not giving an accurate answer to a question”.

tions and rules correctly, they may not trust indirect questioning methods and deliberately answer contrary to fact, i.e., lie, which typically results in false negatives. In particular, respondents carrying a sensitive trait and fearing risk of disclosure when revealing their true status should be encouraged to lie, as admitting to their true status may have negative consequences for them (Krumpal, 2013; Tourangeau & Yan, 2007).

Previous empirical evidence indicates that few respondents report trusting that the ICT, CM or TM actually guarantees anonymity or privacy, with the ICT scoring higher on the trust measure than the CM or TM (Coutts et al., 2011). Some studies also suggest that a substantial share of the respondents seem to believe that indirect question designs involve some kind of a “trick” that may even allow researchers to find out the respondents’ status on the sensitive item, which may foster mistrust (Abernathy, Greenberg, & Horvitz, 1970; I-Cheng, Chow, & Rider, 1972; Soeken & Macready, 1982; Wolter, 2012). Landsheer et al. (1999) demonstrate that respondents who neither understood nor trusted the methods were also less likely to give an honest answer or to answer at all. In addition, Wolter (2012) finds no correlation between understanding and trust as well as understanding and honesty. In line with this finding, Hoffmann et al. (2017) report no correlation between the level of comprehension and the level of perceived privacy protection.

The above discussion suggests that in-depth analysis of respondents’ cognitive processes is required when indirect questioning is applied, in order to make proposals to improve these methods and further encourage respondents to more honest self-reports.

3 The cognitive mechanisms of the survey response

The traditional model of the cognitive mechanisms of the survey response describes four necessary thinking steps in order to provide a correct answer to a given survey question (Tourangeau, 1984; Tourangeau et al., 2000): Respondents need to (1) comprehend the survey question or task at hand; they (2) have to retrieve the relevant information from their memory in order to (3) make a judgement about the question at hand, which they can finally (4) translate into the relevant available response category.

It is commonly accepted that survey questions are inherently subject to misinterpretations by respondents, as a survey situation does not follow a natural conversation, but some kind of standardised, but artificial communication (Mishler, 1986). As a result, it is recommended that survey questions ask about one concept only, use simple language, and are short and precise (e.g., Groves et al., 2009). However, more often than desirable, direct question formats do not follow these rules, let alone more complex indirect questions.

Previous research suggested that the cognitive burden of giving a survey response can also be evaluated according

to how demanding the formal question requirements are (E. Blair & Burton, 1987; Burton & Blair, 1991; Fowler Jr, 1995; Fowler Jr & Mangione, 1990; Hasher & Zacks, 1979, 1984; Mangione, Fowler Jr, & Louis, 1992; Tversky & Kahneman, 1974). For instance, the cognitive burden of answering a question is higher if it asks about difficult content (i.e., the question includes unknown terms or definitions) or entails long instructions and complex sentence structure. Moreover, processing of rules or tasks is more cognitively challenging to survey respondents.⁵ In addition, information that is considered sensitive may evoke emotional responses which are also more challenging to respondents (Bailar, Bailey, & Stevens, 1977; Fellegi, 1964).

Linking these strands of literature and, as a rule of thumb: the more cognitively challenging the formal question characteristics are, the more likely it is that the respondents’ ability to comprehend, retrieve and judge information, and to translate the judgement into a survey answer, is biased.

The ICT, CM, and TM appear to be cognitively more challenging. For instance, these formats ask about several concepts in one question. The ICT asks about various different types of behaviour that are supposedly related, but typically requires thinking about three to five different items at the same time. In addition, one of these types of behaviour is sensitive and may cause emotional reactions. The CM and TM also include at least two concepts, as both questions in the designs need to be unrelated, and one of these concepts captures sensitive information.

While all formats may use simple language, the introductory texts are significantly longer than an average direct question, which is especially relevant for the CM and the TM. Inherently, considering longer text and multiple concepts, these methods are less precise than direct question formats.

In addition, the methods are more complex regarding the information that needs processing. The ICT requires significant memorization along with a mathematical task, which may seem simple, but could potentially cause the respondent stress and result in an incorrect answer. The CM and TM require comprehension and memorization of two unrelated questions, retrieval of the relevant information for both questions and translation into response options that are not visually displayed. Finally, they require a connection of both hypothetical responses into one answer that translates into the given response options.

In sum, survey research theory provides a list of pitfalls for the ICT, CM and TM potentially to cause problems in a survey situation (for an overview, see also Wolter, 2012).

⁵The literature distinguishes between episode enumeration, i.e., recalling episodic events; rule based enumeration, i.e., recalling events applying an occurrence rule or time frame; heuristic based enumeration, i.e., using short cuts to recall information; and automatic enumeration, i.e., using typical behaviour to recall information (see e.g., E. Blair & Burton, 1987; Burton & Blair, 1991)

Paired with previous empirical evidence on the effectiveness of indirect questioning formats, it seems that further research is required to understand fully how respondents perceive these techniques and answer questions based on these methods. This seems to be crucial to draw further conclusions about their effectiveness in (better) eliciting sensitive characteristics and make recommendations to improve the methods.

4 Academic misconduct

Good academic practice is commonly understood as behaviour that complies with the academic code of good practice and the ethical guidelines of academic institutions. However, it is more difficult to define the essence of academic misconduct. Thinking about misconduct more broadly, it includes all behaviours that violate the code of good academic practice, independently of the severity of this breach. The discourse of misconduct demonstrates that it is often linked to the severity, perception, definition, as well as communication of the rules and violations within an academic discipline and to the broader research community (Hesselmann, 2018; Hesselmann, Wienefoet, & Reinhart, 2014). In addition, the visibility of academic misconduct is often relatively low, given that most institutions lack a precise understanding of it (Hesselmann, Graf, Schmidt, & Reinhart, 2017). Only severe and very obvious cases are made public and are well known. In times of increased pressure on academics working in progressively competitive educational systems (e.g., Johann & Neufeld, 2016), academic misconduct is believed to be on the rise.

Given the lack of an overall definition, measuring misconduct is challenging. Often researchers estimate the extent of specific types of misconduct by directly asking survey respondents to provide a self-report of their own malpractices or by directly enquiring about the frequency with which their colleagues and peers engage in unethical behaviour (Daniel, Blount, & Ferrell, 1991; McCabe & Trevino, 1997; Patrzek, Sattler, van Veen, Grunschel, & Fries, 2015). For instance, the German Scientist Survey 2016 reports that one fifth of German academics have been granted authorship without making substantial contributions; around 16% state they insufficiently reviewed the literature in their fields; less than 5% admit to sloppy paper reviews and self-plagiarism; and less than 0.5% indicate that they cooked up their data (Neufeld & Johann, 2016). Another study finds that about one third of academics admit having engaged in questionable research practices, such as falsifying data, plagiarism, or unauthorized use of confidential information, at least once (Martinson, Anderson, & De Vries, 2005).

As uncovering academic misconduct can have severe consequences for researchers – including the loss of reputation for the individual and their institution, such as exclusion from academic institutions and organisations, fines and the deprivation of academic titles – the actual prevalence of aca-

ademic misconduct is likely to be higher than estimated. It is extremely difficult to precisely determine how widespread malpractice actually is by asking about it in surveys (Decoo, 2002; Hesselmann et al., 2014). As such, capturing academic malpractice using self-reports in surveys suffers from under-reporting due to social desirability concerns and the potential risk of disclosure (Krumpal, 2013, 2014; Tourangeau et al., 2000). To illustrate, research matching reports of academic misconduct and researchers' perceptions of misconduct of others indicates that the perception of misconduct of others is about three times higher (17.4%) than the rate of academics admitting misconduct (4.7%) (Gardner, Lidz, & Hartwig, 2005).

We may speculate that even the perceived rates underestimate academic misconduct, as academics may be unaware of others' breaches of research integrity; they have a different understanding of what makes a behaviour unethical; or they have different definitions of malpractice. While the latter are difficult to address, ICT, CM and TM designs may help to tackle the issue of privacy protection and social desirability bias in reporting misconduct.

As academics are presumably very highly educated respondents with the cognitive capacity quickly to process content, they represent a suitable group to study whether ICT, CM and TM are as effective as previously claimed.

5 Method

5.1 Cognitive interviews

To study comprehension and perceived privacy protection of the ICT, CM and TM, we conducted cognitive interviews in order to test whether the techniques work in the context of studies investigating scientific misconduct.⁶ Cognitive interviewing is a commonly used qualitative method to test survey questions and questionnaires, as opposed to survey processes, and pays explicit attention to the respondents' thinking processes (Beatty & Willis, 2007; Collins, 2003, 2015; Willis, 2004). Cognitive interviewing mainly relies on two core techniques: thinking aloud and probing. While the former is used to allow respondents to voice their thoughts about how they have arrived at an answer, the latter allows the interviewer to pose semi-structured, but targeted follow-up questions to clarify terms and uncover potential misunderstandings and/or interpretations (Beatty & Willis, 2007; Willis, 2004). The success of the cognitive interview largely depends on the interaction between the interviewer and the respondent. Respondents may be unwilling or unable to voice their thoughts clearly, while interviewers need to be trained to encourage respondents to elaborate on their answers.

⁶Ethics approval for the project has been obtained by the Sociology Ethics Committee at City, University of London (Soc-REC / 80025566 / 21-03-17)

Table 1
Considered questionable research practices

	Level of severity	
	severe	less severe
Authorship	intentionally using someone else's ideas	claiming first authorship
Data	manipulating empirical data	strategically excluding results
Writing	withholding results	changing hypotheses

5.2 Questionnaire

The questionnaire was divided into two sections: Section A is a short survey questionnaire; Section B is a guide for the interviewers to ask a set of probing questions.

Section A included six questions on more or less severe types of academic misconduct using a three-by-two design, i.e., three question formats asking about two types of academic misconduct each (see Table 1). We ensured that each item was asked by all three methods across the sample by rotating the items in the questionnaire (see Appendix A1).

The ICT was designed as follows: We carefully discussed and pretested the unobtrusive items of the ICT. The aim was to find at least one high prevalence and one low prevalence item along with items that had mid-range prevalences. To build one coherent list, the list items were also required to relate to academic life and practices. A longer list with various items was discussed with researchers unrelated to this project. These colleagues were also asked to provide an estimate of the prevalence of each item. The final two lists of four validated unobtrusive items included statements about commuting more than 50 kilometres to work, frequent participation in research seminars, subscriptions to two academic journals, the use of a Windows computer for work (List A), having lunch with colleagues, presenting work in different languages, taking handwritten notes and having a personal subscription for the print version of at least two newspapers (List B). The respective sensitive item was integrated into the middle of each list. The order of the list items was fixed. Respondents were instructed to indicate the number of items they had “ever” done.

The CM and TM designs rely on pretested instructions provided by Höglinger (2016). We use the traditional birthday question with known probabilities and vary the reference person (father/mother) and the time (birthday/birth month). The hypothetical response option was represented by a binary response code.

The sensitive question on academic misconduct also had a hypothetical binary response-code.

As the CM and TM are very similar methods, we aimed to

avoid respondent fatigue by fixing the order of the questions to separate them a little bit (see Table A1). The questionnaire began with two questions using the CM, followed by two questions asking the *long* list of the ICT, and it closed with two questions using the TM. To prevent item effects, we rotated the item pairs by question format. While our core questionnaire is the same for all respondents, we asked all item pairs across the different question formats. Strictly speaking, we had three short questionnaires, each including six questions with a fixed question format structure, but varying the order of the misconduct items.

To simulate a survey situation, respondents were asked to fill in a paper questionnaire, place it in an envelope, seal it and hand it back to the interviewer at the beginning of the interview. The sealed envelope was passed back to the respondent by the interviewer after the interview. Using the sealed envelope design, we encouraged the respondents to read carefully and answer honestly without the interviewer knowing the answers.

Section B of the study provided follow-up probing questions. This part was structured into two main sub-sections (1) comprehension and (2) honesty.⁷

We distinguished four dimensions of comprehension: Respondents were first asked to repeat the instructions in their own words. In addition, respondents were asked to explain why the questionnaire used unusual question formats and, third, how they thought their answers were better protected by these techniques. Lastly, interviewers also enquired whether respondents had an idea of how their answers could be used for analysis in a meaningful way.

To capture honesty, respondents first had to indicate whether they had answered all questions honestly and, second, give an assessment if other academics would be encouraged to answer honestly. Since the direct question about answering honestly may itself be considered sensitive itself, we believe that asking about other researchers' behaviour is an indirect way of assessing honesty. Finally, respondents were asked to indicate which method offers respondents the highest levels of privacy and anonymity, and about their preferred method.

Both questionnaires were drafted and validated in German and English.

5.3 Respondent selection and recruitment

The sampling strategy for the cognitive interviews was purposively to sample more than 10 and up to 50 respondents, depending on the number and clarity of sampling criteria and attributes (J. Blair & Conrad, 2011; Collins, 2015). We opted for a Parallel Non-Interlocked Quota Sampling Design (see Table 2), which allowed us to specify diversity in

⁷We also include questions evaluating the sensitivity and severity of the respective misconduct. However, this part of the questionnaire is not included in the analysis presented here.

terms of the number of interviews with different types of people within the population (Collins, 2015). In other words, the overall aim was to sample a total of 20 respondents for cognitive interviews along three core characteristics: (1) academic status (PhD level, postdoc/lecturer, senior lecturers/readers/professors); (2) academic field (Natural Sciences, Humanities/Social Science/Business and Economics; Engineering/Computer Science; Law; Health/Medical Sciences) and (3) gender (male, female).

We assume that junior, mid-career and senior academics differ especially in the level of experience and understanding, but also in their perceptions of misconduct. Moreover, previous research has indicated that academic traditions and perceptions vary substantially across disciplines (see e.g., Glänzel, 2002; Johann & Mayer, 2019; Sin, 2011, for discussions on authorship). Thus, it is important broadly to cover the core disciplines as recognised in the Field of Science and Technology (FOS) by the Organisation for Economic Co-operation and Development (OECD, 2007). Finally, gender differences may be observable (Fanelli, Costas, & Larivière, 2015).

We browsed the respective departments at the target institutions in Germany, Switzerland and the United Kingdom to compile a list of prospective participants and substitutes. An initial letter was sent to the prospective respondents, in which we announced that a member of the research team would be in touch by telephone in the next few days (see Appendix B). If contact could not be made within 5 working days, we sent an informal follow-up email. After another 5 working days of non-contact, we sent an advance letter to the substitute respondent on our list. The interviews were pre-arranged at a convenient time and place of their choice without third-party presence.

5.4 Data collection

Cognitive interviewing with academic researchers varying across our three core criteria took place between June and December 2017 in Germany ($n=9$), the United Kingdom ($n=6$) and Switzerland ($n=4$).⁸ All interviews started with a standardised respondent briefing, in which participants were talked through the information sheet (see Appendix C) and asked to provide their written consent to participate in the study as well as their agreement to audio-record the interview (see Appendix D). None of the respondents who had arranged an appointment with us refused to participate. All respondents gave consent to audio-record the interview.

Respondents were first asked to take a few minutes to fill in the paper questionnaire. The interviewer ensured the respondents that they would not be interrupted and no one would interfere with the paper questionnaire. Once the respondent indicated they had answered the six questions, the interviewer handed over a blank envelope in which the respondent was instructed to place the filled-in questionnaire.

The respondent was then asked to seal the envelope and hand it back. We opted for a sealed-envelope approach to ensure the highest possible level of privacy and to encourage honest self-reports. The sealed envelope was handed back to the respondent after the interview. Respondents who did not want to take the envelope with them were asked to destroy it right away.

Within the actual cognitive interview, specific probing questions regarding respondents' understanding of the question formats, honesty of their answers and the perceived sensitivity of the misconduct items were asked. To remind respondents of the different questions, the interviewer presented the participants with show cards that corresponded with the core questionnaire. All interviews were transcribed and anonymised to an extent that the potential risk of disclosure was eliminated.

Analysis strategy. We study the relationship of comprehension, following instructions and honest self-reports. In particular, we aim to disentangle different levels of comprehension beyond understanding of instructions. We propose the following scheme, distinguishing four levels (see Figure 1):

(1) Comprehension of *instructions*, i.e., the respondents' ability to repeat instructions in their own words, as the most basic level. (2) Understanding of the *purpose* of using these methods, i.e., protection of anonymity and privacy. (3) Recognition of *protection*, i.e., acceptance that individual answers are anonymous (i.e., no one can know for sure which answer a respondent has given) and protected. (4) Awareness of *methods of analysis*, i.e., are respondents aware of how their answers can be used for analysis in a meaningful way?

We hypothesise that the first three levels (instructions, purpose, protection) in particular are necessary conditions for respondents to give an honest and correct answer. We are more lenient regarding the methods of analysis, as the analysis strategy is typically beyond the consideration of a survey respondent. It is also excluded from general cognitive models of the survey response and irrelevant during the interview situation. However, having a clear understanding about how individual answers are reflected in the final estimate may increase trust in privacy and anonymity. However, a lack of understanding how answers are used should not considerably affect the honesty of the response.

We assume that, with the increasing cognitive burden, fewer respondents will be able to fit the scheme. Hence, respondents who do not understand the basic instructions will have trouble understanding that these methods help protect their individual anonymity, to recognize that they are protected and to be aware of the analysis strategy.

⁸The final sample encompasses $n=19$ instead of $n=20$ respondents since several attempts to get in contact with our last target respondent as well as the substitutes failed.

Table 2
Matrix of the Parallel Non-Interlocked Quotas

Sampling Criteria	Quota
Field	4 academics from the Natural Sciences 4 academics from Humanities/ Social Sciences/ Business & Economics 4 academics from Engineering/ Computer Science 4 academics from Law 4 academics from the Health/ Medical Sciences
Status	5 PhD candidates 5 postdocs/ lecturers 10 senior lecturers/ readers/ professors
Gender	10 female academics 10 male academics

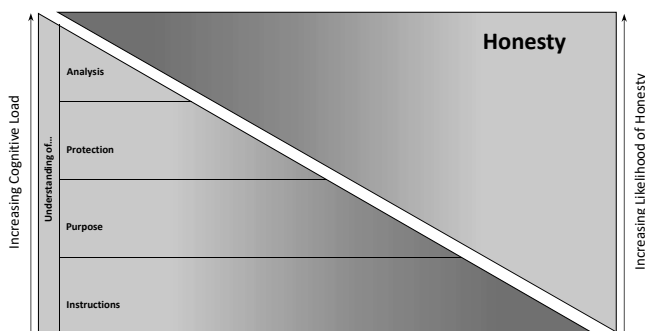


Figure 1. Relationship between comprehension and honesty

Coding. The coding scheme was developed by the core research team. We identified the most important categories for analysis regarding our research question and opted for a five-category coding scheme, four of which referred to different levels of understanding, and one category to honesty, as defined above. Next, we systematically analysed the transcript of all interviews identifying statements relating to each category. The scheme was applied to each individual method. Table 3 summarises the coding scheme and provides relevant examples. Once all interviews had been systematically coded and the results recorded by each individual coder ($n=4$), we evaluated the inter-coder reliability. Few remaining critical cases were discussed with all coders in the team.

6 Results

6.1 Comprehension

Instructions. In line with previous research, comprehension of the instructions for the three methods was very high. Only two respondents did not *clearly* repeat the instructions, yet during the cognitive interview they demonstrated that they had a clear understanding of the instructions. 17 out of 19 respondents were able to recite or explain clearly and correctly the instructions for the three methods.⁹ For exam-

ple, Respondent 4 said about the CM: “[If] one has the same answer to both questions, one has to check A; if one answers differently, one has to check B.”¹⁰ Respondent 8 describes the ICT: “[Here] I had to look at the five bullet points ... and had to assess if this applies to me and if I have done this before and then write down a number [of items] below.”¹¹ About the TM, Respondent 14 said: “[T]his is more similar to parts of the questions one and two ... and really what you’re doing is switching around how you answer them. So ... it’s both with a no or at least one with a yes.” We concluded that there are no observable differences regarding the comprehension of the instructions for the three methods. It is noteworthy that some of the respondents did not see the difference between the CM and TM unless it was specifically pointed out by the interviewer. For instance, when asked about the differences between CM and TM, Respondent 11 replied: “Ah, it’s at least one yes. ... I missed that completely when I read that; ... it is even fuzzier.” However, this should not pose a problem for comparative studies investigating these special questioning methods, as the comparison usually takes place in a between-subject design, where respondents are only presented with one method. This may only be an issue when researchers aim for a within-subject comparison.

Purpose. When asked about the purpose of the unusual question format, 14 out of 19 respondents stated that the purpose was to encourage them to self-report more honestly. Representative is Respondent 2, who stated: “I had the im-

⁹Two respondents used showcards to remind themselves of the instructions.

¹⁰Translated from German: “[W]enn man auf beide Fragen ... die gleiche Antwort hat, muss man A ankreuzen, wenn man beide Fragen unterschiedlich beantwortet, muss man B ankreuzen ...”

¹¹Translated from German: “[Hier] musste ich mir fünf Punkte angucken, ... musste mir überlegen, ob das auf mich zutrifft und ... ich das schon mal gemacht habe und dann eine Zahl hinschreiben”

Table 3
Coding scheme

Code	Description	Examples
(1) Instructions	R able to recite minimum rules correctly	<p>“This is about answering question 1 and then question 2 independently of each other and then to check box A or B, depending on whether one said yes and no, or just yes, or just no.” (CM)</p> <p>“Here it is the sum of yeses that have to be recorded in the end.” (ICT)</p> <p>“Something was changing with the later questions The A was, it could be two yeses or two no’s, but then it changed into two no’s.” (TM)</p>
(2) Purpose ^a	R states that the question format preserves anonymity/privacy	<p>“It sort of gives respondents an opportunity to answer honestly, but in such a way that they don’t feel like they are having to say it directly.”</p> <p>“There is a way of trying to draw out information that perhaps people wouldn’t want to bring to the forefront or may not be comfortable admitting. And the fact [that] that’s anonymous obviously eliminates that to some extent; but even based on that, some people might choose to evade certain answers.”</p>
(3) Protection ^a	R understands how the protection works	<p>“You do not ask direct questions, but one seemingly does not know which answer . . . I have really given to the individual questions, as I have answered several questions at the same time.”</p>
(4) Analysis ^a	R refers to probability methods	<p>“There’s some sort of statistical method that could allow you to collect data across respondents, examine within case variation, and so forth, to get the truth.”</p> <p>“Honestly, no.”</p> <p>“You could do the stats . . . I don’t know the maths . . . , but you could work that out, I suppose.”</p> <p>“Honestly speaking, it was not clear to me what the results could be . . . One could possibly analyse [this] statistically.”</p>
(5) Honesty ^a	R reports that they (a) answered honestly (b) others would, too	<p>“I answered them truthfully . . . I answered it honestly, but I can see there being instances, where there could be a misbalance.”</p> <p>“[I]t’s only their perception of the answer. I suppose perception enters into a lot of these things, because . . . a lot of people . . . do things, potentially, that are in the grey area of research ethics. . . . Some people would lie about these things, because they are never willing to give that information ever”</p>

^a The question was asked about all methods.

pression that this is about anonymity, that I somehow always had another response category, . . . that I had other options that I could admit to without further ado; . . . that I did not necessarily have to reveal my own misconduct.”¹² Another respondent pointed out: “[I]t’s sort of, you’re not forcing people to say which one they are saying yes to, so it’s sort of like a double blind idea of anonymity, because I know from survey questions I have filled out before sometimes the answers can basically give you away.” Most respondents referred to more honest answers due to the indirect way of posing the question or increased anonymity as a consequence of improved protection. Only four respondents did not *clearly* mention privacy concerns or honesty. Two of these mentioned that distraction from the sensitive items might be the

mechanism and, thus, indirectly imply privacy and honesty concerns. Representative is Respondent 6’s answer: “Of course, you somewhat give the impression that you don’t know in the end which answers I really gave to the individual question, as I answered several at the same time.”¹³

¹²Translated from German: “[I]ch hatte schon den Eindruck, dass es um Anonymität geht, dass ich immer irgendwie eine zusätzliche Antwortmöglichkeit hatte, . . . dass ich eben Optionen habe, die ich ohne weiteres zugeben kann; . . . dass ich nicht zwingend mein eigenes Fehlverhalten offenlegen muss.”

¹³Translated from German: “Sie vermitteln natürlich so ein bisschen den Eindruck, dass [Sie] am Ende scheinbar nicht [wissen] welche Antwort ich wirklich gegeben habe auf die einzelnen Fragen, da ich mehrere Fragen gleichzeitig beantwortet habe.”

Protection. Considering the perceived level of protection, 12 out of 19 participants clearly indicated that they felt they were protected, as researchers do not know their individual answers to the different questions. Respondent 14 states: “[A]s long as you know which group I belong to, then you are only going to be able to find out sort of group-level information, not individual-level information.” Regarding the CM and TM, respondents mostly referred to the unknown birthday of their parents. Respondent 1 points out: “If only one question is answered with yes, then it could theoretically also be the birthday and then the probability that I have done it [the misconduct], is higher in your analysis, but you cannot say it for sure ... unless you control for the birthday of my parents.”¹⁴ For the ICT, most respondents suggested that by asking many different questions with an unknown response noise would be introduced, as Respondent 16 mentions: “I guess, because it’s very unimportant whether I use Windows ..., so this mixed up with other questions.”

However, of the remaining seven respondents who did not understand the idea of protection, some feared that it would be somehow possible to figure out their parents’ birthdays. Respondent 11 voiced concern: “Not to come across as incredibly paranoid, but your colleagues might know when your parents’ birthdays are.” For studies that apply CM or TM investigating sensitive topics, such respondent concerns constitute an immediate problem, since respondents consequently will not trust these methods.

Analysis. Only 8 out of 19 respondents have a rough idea how the ICT, CM or TM could be analysed referring to probability theory, aggregation or related key words. Respondent 16 speculated about the CM/TM: “[I]f you were interested in various forms of academic misconduct and you kind of have a normal distribution of birthdates or something, then you could work out what you know which are recurring more frequently than you would expect, something like that. Just because you know about birthdays.” Three respondents directly indicated that they do not have any idea how these data could be analysed in a meaningful way, Respondent 15 clearly states: “I don’t know.” Six respondents referred to some procedures relating to statistics or probabilities without further detail, representative is Respondent 6 saying: “You would need a larger number of data and could possibly analyse this statistically.”¹⁵ The remaining two respondents assumed that researchers could figure out their parents’ birthdays to know about their response to the misconduct item; Respondent 18 presumes: “I envision that this will be overall a very long questionnaire, in which you ask repeatedly about birthdays, fathers and mothers.”¹⁶ Most respondents were only able to answer this question about the CM and TM, as they did not know the short list of the ICT – they were only presented with the long list. Some respondents speculated that we derived the prevalence for the single non-sensitive ICT items beforehand to come up with a distribution that we

plot against their answers. Figure 2b summarizes our results.

In sum, sixteen respondents were able to recite the instructions. 12 out of 17 respondents correctly summarised the instructions and were able to explain the purpose of these methods. In addition, eleven also gave an appropriate explanation of how their answers were protected. Apart from the instructions, the purpose and the protection mechanism, only seven respondents were aware of how their answers could be analysed in a meaningful way. These results also support our assumption that the different levels of understanding are increasing in cognitive demand: with an increasing level of understanding, more and more respondents drop out. Only one respondent deviated from this pattern by failing in an easy task (*purpose*), but succeeding in an arguably more difficult task (*protection*).

6.2 Honesty

Looking at whether respondents would answer truthfully, 17 out of 19 respondents stated that they had answered honestly. Nine respondents were also confident that the vast majority of researchers would give honest answers when asked about academic misconduct using these question formats. However, eight respondents voiced serious doubts about asking survey questions in this way. For example, Respondent 11 raises the question: “[i]f they are willing to fudge data, who is going to say that they won’t do it again?” One respondent even admitted that they had answered “almost all” questions truthfully, indicating that they had in fact not answered all questions honestly. The same person still believed other researchers would answer all questions truthfully.

6.3 Comprehension and honesty

One core objective of this study was to investigate the link between comprehension and honesty. We expected that the better the respondents’ comprehension of the methods was, the higher the degree of honesty would be. However, the results from the cognitive interviews do not seem to support this notion.

Figure 2b plots the share of respondents stating they answered honestly *and* other researchers would answer honestly against their respective level of comprehension. The

¹⁴Translated from German: “[W]enn mindestens eine Frage, die mit ja beantwortet ist, dann kann es zumindest ja theoretisch auch der Geburtstag sein und damit ist die Wahrscheinlichkeit, dass ich es schon mal gemacht habe, natürlich erhöht in Ihrer Auswertung, aber Sie können es nicht sicher sagen ... es sei denn Sie kontrollieren den Geburtstag meiner Eltern.”

¹⁵Translated from German: “Sie müssen eine große Menge von Daten haben [u]nd dann können Sie möglicherweise statistisch was auswerten.”

¹⁶Translated from German: “Ich stelle mir vor, das wird dann insgesamt ein sehr langer Fragebogen, wo immer wieder nach den Geburtstagen, immer wieder nach Vater, Mutter gefragt wird.”

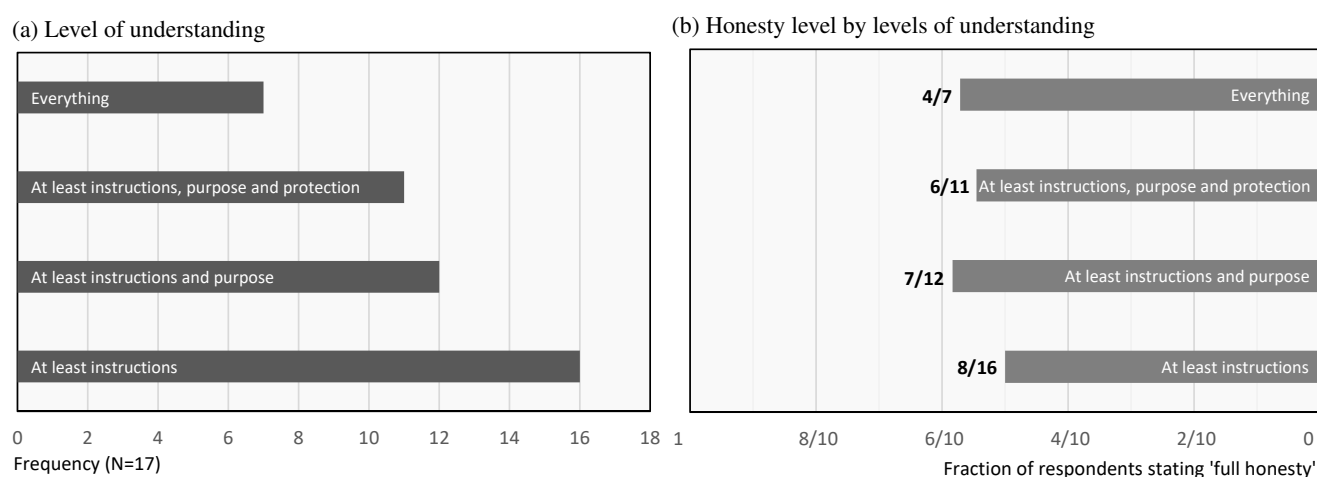


Figure 2. Relation between comprehension and honesty. Panel (a) plots conditional frequencies of increasing levels of comprehension of the anonymisation technique (easiest level at the bottom; hardest level on top). Panel (b) plots the conditional fraction of respondents who stated they had answered honestly and indicated that others would respond truthfully, too ('full honesty'), among those who achieved the respective level of understanding. To understand better how the two charts interrelate, see the following example: Out of 16 respondents who understood the instructions, 12 also understood the purpose (second bar from bottom, Panel (a)); of these 12 respondents, seven stated 'full honesty' (second bar from bottom, Panel (b)). The sample size is reduced to 17 because in two cases either non-response for understanding or honesty occurred.

figure indicates that 8 out of 16 respondents understood the instructions and were fully confident that people including themselves would answer truthfully. 7 out of 12 who comprehended the instructions and the purpose said people including themselves would be fully honest about misconduct when asked indirectly. This also holds for 6 out of 11 who fully understood the protection mechanism. Only 4 out of 7 who displayed awareness of how to analyse the data said academics would answer these indirect questions on academic misconduct honestly. Overall, honesty does not seem to vary with increasing levels of comprehension.

7 Discussion

Indirect questioning formats have a long tradition and reputation to estimate sensitive behaviour better. They are designed to ask about the sensitive characteristic using unusual question formats and/or auxiliary aids to obscure the meaning of response categories. Over the past decade, the ICT, CM and TM were praised as successful and effective methods to reduce social desirability bias and to protect respondents' privacy and anonymity in order to encourage more honest self-reports. However, recent research has voiced criticism about the effectiveness of these methods, especially when studies uncovered the occurrence of false negatives, false positives and other design issues, resulting in biased estimates (Coutts et al., 2011; Droitcour et al., 1991; Höglinger & Diekmann, 2017; Höglinger & Jann, 2018; Landsheer et al., 1999; Wolter, 2012; Wolter & Laier, 2014). One major

concern is that respondents struggle to comprehend the instructions, to follow the rules correctly and, thus, to answer incorrectly.

We take a qualitative approach to study the respondents' cognitive processes when answering questions using the ICT, CM and TM. Employing cognitive interviews with academics allows us to study a very particular and highly educated group. We assumed that, if academics are not capable of fully comprehending the instructions, the methods are likely to fail on samples of general populations (Hoffmann et al., 2017; Schnell et al., 2019; Wolter, 2012).

While most of the respondents understand the instructions and are able to repeat them in their own words, fewer are able to recognise that the format is meant to protect their privacy and anonymity better, and feel encouraged to answer honestly. An even smaller number of respondents is aware of the way the data can be analysed. When asked about honesty, the interviews revealed that people are not necessarily encouraged to answer honestly and have doubts about others answering truthfully: Only half of the sample expect the methods to yield truthful answers. In addition, comprehension of the formats and honesty are seemingly unrelated.

Overall, our results indicate that the methods may be too sophisticated even for highly educated survey respondents. Embedding our findings in the survey methodology theory, especially the cognitive mechanism model of the survey response and theory on the effect of survey question characteristics, it appears that these questions are more problematic

than previously claimed. The formats do not consider basic rules of question design asking lengthy and imprecise questions about more than one theoretical concept. They also increase the cognitive burden on respondents requiring complex memorization and mathematical tasks, which may appear simple, but may evoke stress and rejection by respondents. Accordingly, our results support doubts that the ICT as well as RRT produce more valid results than direct question formats (Wolter, 2012).

8 Conclusion

An increasing interest in sensitive types of behaviour and attitudes in surveys has led to a growing literature on how to elicit this information more accurately. One dominant line of research is to protect respondents through indirect questioning, often tied to including a random element. However, such an approach requires respondents to focus and spend time and efforts on the task at hand. Ideally, it also requires statistical knowledge and understanding. Psychological research suggests that ordinary citizens, but even applied quantitative researchers, have problems understanding and processing statistical concepts (Cohen, 1994; Gigerenzer & Todd, 2000). Thus, it remains unclear to what extent these methodological approaches help elicit sensitive answers from survey respondents.

Our research design, employing cognitive interviews that are qualitative in nature, demonstrated that even highly educated respondents have difficulties comprehending how the ICT, CM and TM protect sensitive answers. The findings suggest that we should be much more wary about the use and the performance of these techniques. This raises concerns whether the methods actually fulfil their purpose of encouraging more honest self-reports when applied to general population samples. If highly educated people find these questions too cognitively challenging, we may speculate that less well-educated respondents struggle even more. Hence, at least awareness for this issue needs to be created regarding the applicability of the methods when general populations are concerned (see also Schnell et al., 2019, for a similar empirical account).

Of course, our research has limitations. We provided a small n qualitative account of the cognitive processes regarding these special question formats on a highly educated sample. It would be interesting to repeat this design and include a wider audience that would allow us to draw conclusions about the cognitive processes of less educated respondents as well. Furthermore, conclusions drawn on the basis of small n qualitative studies may lack external validity. While our findings may provide some additional guidelines regarding design considerations for ICT, CM and TM applications, it is up to future research further to study the effectiveness of these question formats with special focus on the relationship between comprehension, trust and response behaviour.

Applied researchers should be advised to aim for an operationalisation that is easier to process for respondents instead of designing question formats that require sophisticated tasks. As such, the straightforward ICT design appears to be less challenging for respondents, because it generally involves shorter instructions and a simple counting task. By contrast, CM/TM instructions are usually longer and the commonly used birthday question may evoke mistrust. Thinking of possible alternatives, the birthday question may be replaced by an unobtrusive item that is contextually related to the sensitive question and whose prevalence can be estimated in a different subsample. For example, comparative designs that directly ask about the sensitive item in one subsample and in another subsample measure the prevalence using a CM/TM format would allow us to determine the prevalence of the unobtrusive item for the CM/TM by including it in the subsample, using direct questioning techniques.

Acknowledgements

We thank the anonymous reviewers for their concise comments and constructive feedback on our work. We also thank Nicola Babamov, Christophe Heger, Beatrice Schulz, Madeleine Siegel, Nikita Sorgatz, and Anastasiia Tcypina for their assistance with the transcription of the interviews as well as Brian Cooper for proofreading. David Johann would like to thank the German Center for higher Education Research and Science (DZHW) for their support in providing fund for interviewing and transcription. Kathrin Thomas was Research Associate at City, University of London, when the initial phase of the project started and would like to thank City for its support.

Heiko Rauhut acknowledges funding from the SNSF by the Starting Grant BSS-GIO 155981 “Social norms, cooperation and conflict in scientific collaborations”.

Authors in alphabetical order. Idea: David Johann. Design: Julia Jerke, David Johann, Heiko Rauhut, Kathrin Thomas. Interviews: Julia Jerke, David Johann, Kathrin Thomas. Coding and Analysis: Julia Jerke, David Johann, Heiko Rauhut, Kathrin Thomas. Text: Julia Jerke, David Johann, Heiko Rauhut, Kathrin Thomas. Final approval of the paper: Julia Jerke, David Johann, Heiko Rauhut, Kathrin Thomas.

References

- Abernathy, J. R., Greenberg, B. G., & Horvitz, D. G. (1970). Estimates of induced abortion in urban north carolina. *Demography*, 7(1), 19–29. doi:10.2307/2060019
- Bailar, B., Bailey, L., & Stevens, J. (1977). Measures of interviewer bias and variance. *Journal of Marketing Research*, 14(3), 337–343.

- Beatty, P. C. & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287–311.
- Biemer, P. & Brown, G. (2005). Model-based estimation of drug use prevalence using item count data. *Journal of Official Statistics*, 21(2), 287–308.
- Blair, E. & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, 14(2), 280–288.
- Blair, G. & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47–77.
- Blair, G., Imai, K., & Park, B. (2010). Statistical methods for the item count technique and list experiment. The Comprehensive R Archive Network (CRAN). Retrieved from <http://CRAN.R-project.org/package=list>
- Blair, J. & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75(4), 636–658.
- Boruch, R. F. (1971). Assuring confidentiality of responses in social research: A note on strategies. *The American Sociologist*, 6(4), 308–311.
- Burton, S. & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly*, 55(1), 50–79.
- Campbell, A. A. (1987). Randomized response technique. *Science*, 236(4805), 1049–1049. doi:10.1126/science.3576215
- Clark, S. J. & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods*, 3(2), 160–168.
- Cohen, J. (1994). The earth is round ($p < .05$). *American psychologist*, 49(12), 997–1003.
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 229–238.
- Collins, D. (2015). *Cognitive interviewing practice*. London: Sage.
- Comşa, M. & Postelnicu, C. (2012). Measuring social desirability effects on self-reported turnout using the item count technique. *International Journal of Public Opinion Research*, 25(2), 153–172.
- Coutts, E., Jann, B., Krumpal, I., & Näher, A.-F. (2011). Plagiarism in student papers: Prevalence estimates using special techniques for sensitive questions. *Jahrbücher für Nationalökonomie und Statistik*, 231(5-6), 749–760.
- Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. *Personnel Psychology*, 47(4), 817–829.
- Daniel, L. G., Blount, K. D., & Ferrell, C. M. (1991). Academic misconduct among teacher education students: A descriptive-correlational study. *Research in Higher Education*, 32(6), 703–724.
- De Jonge, C. P. K. & Nickerson, D. W. (2014). Artificial inflation or deflation? Assessing the item count technique in comparative surveys. *Political Behavior*, 36(3), 659–682.
- Decoo, W. (2002). *Crisis on campus: Confronting academic misconduct*. Cambridge, MA: The MIT Press.
- Díaz-Cayeros, A., Magaloni, B., Matanock, A., & Romero, V. (2011). *Living in fear: Mapping the social embeddedness of drug gangs and violence in Mexico*. Manuscript: University of California at San Diego.
- Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 185–210). New York, NY: John Wiley & Sons.
- Edgell, S. E., Duchan, K. L., & Himmelfarb, S. (1992). An empirical test of the unrelated question randomized response technique. *Bulletin of the Psychonomic Society*, 30(2), 153–156. doi:10.3758/BF03330425
- Edgell, S. E., Himmelfarb, S., & Duchan, K. L. (1982). Validity of forced responses in a randomized response model. *Sociological Methods & Research*, 11(1), 89–100. doi:10.1177/0049124182011001005
- Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS One*, 10(6), e0127556.
- Fellegi, I. P. (1964). Response variance and its estimation. *Journal of the American Statistical Association*, 59(308), 1016–1041.
- Fowler Jr, F. J. (1995). *Improving survey questions: Design and evaluation*. London: Sage.
- Fowler Jr, F. J. & Cosenza, C. (2009). Design and evaluation of survey questions. In L. Bickman & D. J. Rog (Eds.), *The sage handbook of applied social research methods* (pp. 375–412). London: Sage.
- Fowler Jr, F. J. & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. London: Sage.
- Fox, J. A. & Tracy, P. E. (1980). The randomized response approach: Applicability to criminal justice research and evaluation. *Evaluation Review*, 4(5), 601–622.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505.

- Gardner, W., Lidz, C. W., & Hartwig, K. C. (2005). Authors' reports about research integrity problems in clinical trials. *Contemporary Clinical Trials*, 26(2), 244–251.
- Gigerenzer, G. & Todd, P. M. (2000). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Gilens, M., Sniderman, P. M., & Kuklinski, J. H. (1998). Affirmative action and the politics of realignment. *British Journal of Political Science*, 28(1), 159–183.
- Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, 50(3), 461–473.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77(S1), 159–172.
- Grant, J. T., Moon, R., & Gleason, S. (2012). Asking many, many sensitive questions: “person-count” method for social desirability bias. Paper presented on the MAPOR Conference, 2012. Retrieved from <https://docplayer.net/35140037-Asking-many-many-sensitive-questions-person-count-method-for-social-desirability-bias.html>
- Grant, J. T., Moon, R., & Gleason, S. (2014). Asking many, many sensitive questions: “person-count” method for social desirability bias. *Unpublished Manuscript*, 1–23.
- Greenberg, B. G., Abul-El, A.-L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64(326), 520–539.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: John Wiley & Sons.
- Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of experimental Psychology*, 108(3), 356–388.
- Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, 39(12), 1372–1388.
- Hesselmann, F. (2018). Science and its others: Examining the discourse about scientific misconduct through a post-colonial lens. *Identities*, 26(4), 1–19.
- Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology*, 65(6), 814–845.
- Hesselmann, F., Wienefoet, V., & Reinhart, M. (2014). Measuring scientific misconduct. lessons from criminology. *Publications*, 2(3), 61–70.
- Hoffmann, A., de Puiseau, B. W., Schmidt, A. F., & Musch, J. (2017). On the comprehensibility and perceived privacy protection of indirect questioning techniques. *Behavior Research Methods*, 49(4), 1470–1483.
- Hoffmann, A., Diedenhofen, B., Verschuere, B., & Musch, J. (2015). A strong validation of the Crosswise Model using experimentally-induced cheating behavior. *Experimental Psychology*, 62(6), 403–414.
- Hoffmann, A. & Musch, J. (2016). Assessing the validity of two indirect questioning techniques: A stochastic lie detector versus the crosswise model. *Behavior Research Methods*, 48(3), 1032–1046.
- Höglinger, M. (2016). Revealing the truth? Validating the randomized response technique for surveying sensitive topics. *PhD Dissertation ETH Zürich*, <https://www.research-collection.ethz.ch/handle/20.500.11850/115919>.
- Höglinger, M. & Diekmann, A. (2017). Uncovering a blind spot in sensitive question research: False positives undermine the crosswise-model RRT. *Political Analysis*, 25(1), 131–137.
- Höglinger, M. & Jann, B. (2018). More is not always better: An experimental individual-level validation of the randomized response technique and the crosswise model. *PLoS ONE*, 13(8), e0201770.
- Höglinger, M., Jann, B., & Diekmann, A. (2016). Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model. *Survey Research Methods*, 10(3), 171–187.
- Holbrook, A. L. & Krosnick, J. A. (2010). Measuring voter turnout by using the randomized response technique: Evidence calling into question the methods validity. *Public Opinion Quarterly*, 74(2), 328–343.
- Horvitz, D. G., Simmons, W. R., & Shah, B. (1968). Unrelated question randomized response model. *Journal of the American Statistical Association*, 63(322), 754–754.
- I-Cheng, C., Chow, L. P., & Rider, R. V. (1972). The randomized response technique as used in the taiwan outcome of pregnancy study. *Studies in Family Planning*, 3(11), 265–269.
- Jann, B., Jerke, J., & Krumpal, I. (2012). Asking sensitive questions using the crosswise model: An experimental survey measuring plagiarism. *Public Opinion Quarterly*, 76(1), 32–49.
- Jerke, J. & Krumpal, I. (2013). Plagiate in studentischen arbeiten. *Methoden, Daten, Analysen*, 7(3), 347–368.
- Johann, D. & Mayer, S. J. (2019). The perception of scientific authorship across domains. *Minerva*, 57(2), 175–196. doi:10.1007/s11024-018-9363-3
- Johann, D. & Neufeld, J. (2016). *Performance-based allocation of funds, pressure to publish and publication strategies*. Paper presented at the 4S/EASST Conference 2016 (Barcelona, 31 August – 3 September).

- Johann, D. & Thomas, K. (2017). Testing the validity of the crosswise model: A study on attitudes towards Muslims: A validation study. *Survey Methods: Insights from the Field*. doi:10.13094/SMIF-2017-00001
- Kirchner, A. (2015). Validating sensitive questions: A comparison of survey and register data. *Journal of Official Statistics*, 31(1), 31–59.
- Korndörfer, M., Krumpal, I., & Schmukle, S. C. (2014). Measuring and explaining tax evasion: Improving self-reports using the crosswise model. *Journal of Economic Psychology*, 45, 18–32.
- Krumpal, I. (2012). Estimating the prevalence of xenophobia and antisemitism in Germany: A comparison of randomized response and direct questioning. *Social Science Research*, 41(6), 1387–1403.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047.
- Krumpal, I. (2014). Social desirability bias and context in sensitive surveys. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 6037–6043). Dordrecht.
- Krumpal, I., Jann, B., Auspurg, K., & von Hermann, H. (2015). Asking sensitive questions: A critical account of the randomized response technique and related methods. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving survey methods: Lessons from recent research* (pp. 122–136). New York, NY: Routledge.
- Kuk, A. Y. (1990). Asking sensitive questions indirectly. *Biometrika*, 77(2), 436–438.
- Kundt, T. (2014). *Applying benford's law to the crosswise model: Findings from an online survey on tax evasion*. Working Paper 148, Helmut Schmidt University Hamburg.
- Kundt, T., Misch, F., & Nerré, Birger. (2017). Re-assessing the merits of measuring tax evasion through business surveys: An application of the crosswise model. *International Tax and Public Finance*, 24(1), 112–133.
- LaBrie, J. W. & Earleywine, M. (2000). Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique. *Journal of Sex Research*, 37(4), 321–326.
- Landsheer, J. A., Van Der Heijden, P., & Van Gils, G. (1999). Trust and understanding, two psychological aspects of randomized response. *Quality and Quantity*, 33(1), 1–12.
- Lensvelt-Mulders, G. J. & Boeije, H. R. (2007). Evaluating compliance with a computer assisted randomized response technique: A qualitative study into the origins of lying and cheating. *Computers in Human Behavior*, 23(1), 591–608.
- Lensvelt-Mulders, G. J., Hox, J. J., Van der Heijden, P. G., & Maas, C. J. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research*, 33(3), 319–348.
- Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, 71(354), 269–275. doi:10.1080/01621459.1976.10480332
- Mangione, T. W., Fowler Jr, F. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, 8(3), 293–307.
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737–738.
- McCabe, D. L. & Trevino, L. K. (1997). Individual and contextual influences on academic dishonesty: A multi-campus investigation. *Research in Higher Education*, 38(3), 379–396.
- Miller, J. D. (1984). *A new survey technique for studying deviant behavior*. Ph.D Thesis, George Washington University, Department of Sociology, Washington, DC.
- Mishler, E. G. (1986). *Research interviewing. Context and narrative*. Cambridge: Harvard University Press.
- Moshagen, M., Hilbig, B. E., Erdfelder, E., & Moritz, A. (2014). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology*, 61(1), 48–54.
- Nakhaee, M. R., Pakravan, F., & Nakhaee, N. (2013). Prevalence of use of anabolic steroids by bodybuilders using three methods in a city of Iran. *Addiction & Health*, 5(3–4), 77–82.
- Neufeld, J. & Johann, D. (2016). *Wissenschaftlerbefragung 2016 Variablenbericht – Häufigkeitsauszählung*. Hannover/Berlin: DZHW.
- OECD. (2007). *Working party of national experts on science and technology indicators. revised field of science and technology (fos) classification in the frascati manual*. Paris: OECD.
- Ostapczuk, M., Moshagen, M., Zhao, Z., & Musch, J. (2009). Assessing sensitive attributes using the randomized response technique: Evidence for the importance of response symmetry. *Journal of Educational and Behavioral Statistics*, 34(2), 267–287.
- Patrzek, J., Sattler, S., van Veen, F., Grunschel, C., & Fries, S. (2015). Investigating the effect of academic procrastination on the frequency and variety of academic misconduct: A panel study. *Studies in Higher Education*, 40(6), 1014–1029.
- Rayburn, N. R., Earleywine, M., & Davison, G. C. (2003). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment & Trauma*, 6(2), 137–152.

- Schnell, R., Thomas, K., & Noack, M. (2019). *Do respondent education and income affect survey estimates based on the crosswise model?* Unpublished manuscript.
- Shamsipour, M., Yunesian, M., Fotouhi, A., Jann, B., Rahimi-Movaghar, A., Asghari, F., & Akhlaghi, A. A. (2014). Estimating the prevalence of illicit drug use among students using the crosswise model. *Substance Use & Misuse*, 49(10), 1303–1310.
- Sin, S.-C. J. (2011). International coauthorship and citation impact: A bibliometric study of six LIS journals, 1980–2008. *Journal of the Association for Information Science and Technology*, 62(9), 1770–1783.
- Soeken, K. L. & Macready, G. B. (1982). Respondents' perceived protection when using randomized response. *Psychological Bulletin*, 92(2), 487–489.
- Thomas, K., Johann, D., Kritzing, S., Plescia, C., & Zeglovits, E. (2016). Estimating sensitive behavior: The ICT and high-incidence electoral behavior. *International Journal of Public Opinion Research*, 29(1), 157–171.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines. report of the advanced research seminar on cognitive aspects of survey methodology* (pp. 73–100). Washington: National Academies Press.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.
- Tsuchiya, T., Hirai, Y., & Ono, S. (2007). A study of the properties of the item count technique. *Public Opinion Quarterly*, 71(2), 253–272.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking sensitive questions: A statistical power analysis of randomized response models. *Psychological methods*, 17(4), 623–641.
- Umesh, U. N. & Peterson, R. A. (1991). A critical evaluation of the Randomized Response method applications, validation, and research agenda. *Sociological Methods & Research*, 20(1), 104–138.
- Vakilian, K., Abbas Mousavi, S., Keramat, A., & Chaman, R. (2016). Knowledge, attitude, self-efficacy and estimation of frequency of condom use among Iranian students based on a Crosswise Model. *International Journal of Adolescent Medicine and Health*, 30(1), 20160010.
- Vakilian, K., Mousavi, S. A., & Keramat, A. (2014). Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a Crosswise Model study. *BMC Research Notes*, 7(28), 1–4.
- Van Der Heijden, P. G., Gils, G. v., Bouts, J., & Hox, J. (2000). A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research*, 28(4), 505–537.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63–69.
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. London: Sage.
- Wolter, F. (2012). *Heikle Fragen in Interviews: Eine theoretische und empirische Validierung der Randomized Response-Technik*. Wiesbaden: VS Verlag.
- Wolter, F. & Laier, B. (2014). The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. *Survey Research Methods*, 8(3), 153–168.
- Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: Design and analysis. *Metrika*, 67(3), 251–263.

Appendix A
Table

Table A1

Appendix A: Item rotation

Questionnaire ID	Level of Severity	CM	ICT	TM
Q1	less severe	claiming first authorship	strategically excluding analyses	changing hypotheses
	severe	intentionally using someone else's ideas	intentionally manipulating empirical data	withholding results
Q2	less severe	changing hypotheses	claiming first authorship	strategically excluding analyses
	severe	withholding results	intentionally using someone else's ideas	intentionally manipulating empirical data
Q3	less severe	strategically excluding analyses	changing hypotheses	claiming first authorship
	severe	intentionally manipulating empirical data	withholding results	intentionally using someone else's ideas

Appendix B
Questionnaire (German)

Vorstellung

Guten Morgen/Nachmittag/Abend. Mein Name ist ... und ich bin vom Vielen Dank, dass Sie zugestimmt haben, an dieser Studie teilzunehmen.

Ich würde gerne damit beginnen, das Prozedere etwas genauer zu erklären. Ich werde Ihnen zunächst einen Papierfragebogen aushändigen und Sie darum bitten, diesen Fragebogen auszufüllen. Der Fragebogen beinhaltet vergleichsweise neuartige Befragungstechniken, die Ihre Privatsphäre schützen sollen. Lesen Sie sich bitte zunächst die Instruktionen genau durch und beantworten Sie dann die Fragen. Nachdem Sie damit fertig sind, würde ich Sie bitten, den Fragebogen in diesen Umschlag zu stecken und sicher zu verschließen.

Daran anschließend werde ich Ihnen einige Nachfragen zu Ihrem Verständnis der Fragen und Methoden stellen. Dabei geht es mir vor allem darum zu erfahren, wie Sie zu Ihren Antworten gekommen sind und ob Sie Bedenken beim Beantworten der Fragen hatten. Für den Erfolg der Studie ist es wichtig, dass Sie mir dabei ehrlich antworten.

Sie finden alle Informationen zu dieser Studie auf diesem Informationsblatt [Informationsblatt aushändigen]. Da es sich bei diesem Projekt um eine wissenschaftliche Studie handelt, muss ich Sie um eine schriftliche Einverständniserklärung bitten, dass Sie an dieser Studie teilnehmen möchten. Wenn Sie damit einverstanden sind, würde ich Sie bitten, dies durch eine Unterschrift auf diesem Formular zu bestätigen [Einverständniserklärung vorlegen]. Dieses Formular dient als Nachweis für die Ethikkommission.

Selbstverständlich ist die Befragung und das anschließende Gespräch vertraulich. Um alle Information möglichst genau dokumentieren zu können, würde ich dieses Gespräch aber gerne aufzeichnen. Wenn Sie damit einverstanden sind, würde ich sie abermals bitten, mir dies mit Ihrer Unterschrift zu bestätigen [Einverständniserklärung vorlegen].

Vielen Dank, dann kann es losgehen.

Q1.

Bei dieser Frage werden zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit Ja oder mit Nein), schreiben Sie dies aber nicht auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar nach folgenden Regeln:

Lautet Ihre Antwort auf beide Fragen Nein oder auf beide Fragen Ja, so setzen Sie Ihr Kreuz bei (A). Lautet Ihre Antwort auf eine der Fragen Ja und auf eine Nein, so setzen Sie Ihr Kreuz bei (B).

Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne.

- Ist der Geburtstag Ihrer Mutter im Januar oder Februar?
- Haben Sie schon mal die Erstautorenschaft für ein Papier beansprucht, obwohl Sie wussten, dass jemand anderes mehr beigetragen hat als Sie selbst?

Wie lauten die Antworten auf beide Fragen?

- ☐ Auf beide Fragen Ja oder auf beide Fragen Nein
- ☐ Auf eine der Fragen Ja und auf eine Nein

Q2.

Bei dieser Frage werden wieder zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte abermals zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit Ja oder mit Nein), schreiben Sie dies aber nicht auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar nach folgenden Regeln:

Lautet Ihre Antwort auf beide Fragen Nein oder auf beide Fragen Ja, so setzen Sie Ihr Kreuz bei (A). Lautet Ihre Antwort auf eine der Fragen Ja und auf eine Nein, so setzen Sie Ihr Kreuz bei (B).

Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne.

- Ist der Geburtstag Ihres Vaters im Januar oder Februar?
- Haben Sie schon mal vorsätzlich die Ideen eines anderen Wissenschaftlers/einer anderen Wissenschaftlerin verwendet, ohne dies kenntlich zu machen?

Wie lauten die Antworten auf beide Fragen?

- ☐ Auf beide Fragen Ja oder auf beide Fragen Nein
- ☐ Auf eine der Fragen Ja und auf eine Nein

Q3.

Untenstehend finden Sie eine Liste mit Aktivitäten, die manche Wissenschaftler tun, andere aber nicht. Bitte sagen Sie mir, *wie viele* dieser Dinge Sie schon mal gemacht haben. Bitte sagen Sie mir nicht, welche dieser Aktivitäten Sie schon mal gemacht haben, sondern *wie viele*:

- An einem typischen Arbeitstag fahre ich in eine Richtung mehr als 50 Kilometer zur Arbeit.
- Ich nehme regelmäßig an Forschungstreffen meiner Institution teil.
- Ich habe schon mal strategisch Analysen aus meinem Papier weggelassen, um den Publikationserfolg zu erhöhen.
- Ich habe ein privates Abo für mindestens zwei akademische Journale.
- Ich benutze vor allem Windows für meine Arbeit.

Bitte tragen Sie unten ein *wie viele* dieser Aktivitäten Sie schon mal gemacht haben.

Q4.

Untenstehend finden Sie wieder eine Liste mit Aktivitäten, die manche Wissenschaftler tun, andere aber nicht. Bitte sagen Sie mir, *wie viele* dieser Dinge Sie schon mal gemacht haben. Bitte sagen Sie mir nicht, welche dieser Aktivitäten Sie schon mal gemacht haben, sondern *wie viele*:

- Ich treffe mich jede Woche mit Kollegen zum Mittagessen.
- In den letzten 12 Monaten habe ich meine Arbeit in mehr als 3 Sprachen präsentiert.
- Ich habe schon mal vorsätzlich empirische Daten manipuliert, um meine Forschungsfrage zu bestätigen.
- Ich mache regelmässig handschriftliche Notizen für wissenschaftliche Zwecke.
- Ich habe ein privates Abo für die Printversionen von mindestens zwei Zeitungen.

Bitte tragen Sie unten ein *wie viele* dieser Aktivitäten Sie schon mal gemacht haben.

Q5.

Bei dieser Frage werden wieder zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte wieder zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit Ja oder mit Nein), schreiben Sie dies aber nicht auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar diesmal nach folgenden Regeln:

Lautet Ihre Antwort auf beide Fragen Nein, so setzen Sie Ihr Kreuz bei (A).

Lautet Ihre Antwort auf mindestens eine der Fragen Ja, so setzen Sie Ihr Kreuz bei (B).

Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne.

- Ist der Geburtstag Ihres Vaters am 1., 2., 3., 4. oder 5. Tag des Monats?
- Haben Sie schon mal eine Hypothese geändert, nachdem Sie die empirischen Ergebnisse kannten, um den Eindruck zu vermitteln, dass sie sich bestätigt?

Wie lauten die Antworten auf beide Fragen?

☐ Auf beide Fragen Nein

☐ Auf mindestens eine Frage Ja

Q6.

Und noch mal. Bei dieser Frage werden wieder zwei Fragen in einem Block gestellt. Überlegen Sie sich bitte wieder zuerst, wie Sie die beiden Fragen einzeln beantworten würden (entweder mit Ja oder mit Nein), schreiben Sie dies aber nicht auf! Je nachdem wie Ihre Antworten auf diese beiden Fragen lauten, kreuzen Sie im Anschluss bitte Möglichkeit (A) oder (B) an, und zwar diesmal nach folgenden Regeln:

Lautet Ihre Antwort auf beide Fragen Nein, so setzen Sie Ihr Kreuz bei (A).

Lautet Ihre Antwort auf mindestens eine der Fragen Ja, so setzen Sie Ihr Kreuz bei (B).

Ihre Privatsphäre bleibt geschützt, da ich Ihre Antworten auf die einzelnen Fragen nicht kenne.

- Ist der Geburtstag Ihrer Mutter am 1., 2., 3., 4. oder 5. Tag des Monats?
- Haben Sie schon mal Ergebnisse zurückgehalten, weil diese im Widerspruch zu Ihrer vorherigen Forschung standen?

Wie lauten die Antworten auf beide Fragen?

☐ Auf beide Fragen Nein

☐ Auf mindestens eine Frage Ja

Könnten Sie bitte den Papierfragebogen in diesen Umschlag geben [Umschlag aushändigen] und diesen sicher verschließen.

Probes

Dies war eine ungewöhnliche Art von Umfrage. Ich werde Ihnen nun diese Fragen noch mal vorlegen [Showcard vorlegen]. Ich möchte nochmal betonen, dass ich kein Interesse daran habe, wie Sie geantwortet haben, sondern an was Sie gedacht haben als Sie die Fragen gehört/gelesen haben. Bitte denken Sie einfach laut nach wenn Sie mir die nächsten Fragen beantworten.

Verständnis.

- *Was denken Sie, warum haben wir Ihnen diese Fragen so gestellt?*
- *Bitte sehen Sie sich die drei Frageformate noch einmal an. Was haben Sie gedacht als Sie diese Fragen beantwortet haben? Wie sind Sie auf eine Antwort gekommen?*
- *Hatten Sie irgendwelche Probleme mit den Anweisungen? [Haben Sie irgendwelche Anregungen zur Verbesserung?]*
- *Könnten Sie die Anweisungen noch mal mit eigenen Worten wiederholen?*
- *Könnten Sie mir erklären warum Ihre Antworten mit diesen Methoden besser geschützt sind?*
- *Was denken Sie, wie können Wissenschaftler Ihre Antworten sinnvoll zur Datenanalyse verwenden?*
- *[Für jede Methode separat fragen:] Denken Sie, dass es eine Antwort gibt, die es niemandem erlaubt herauszufinden, dass Sie [sensitives Item einfügen] gemacht haben?*

Ehrlichkeit.

- *Denken Sie andere Wissenschaftler würden ehrlich antworten?*
- *Haben Sie ehrlich geantwortet?*
- *Was denken Sie, welche der Methoden sichert Ihre Privatsphäre am besten?*
- *Und wie ist das mit den anderen Methoden? Gibt es eine Methode, die Ihnen eher unangenehm ist?*
- *Haben Sie irgendetwas in den Anweisungen vermisst, dass Ihnen geholfen hätte Ihre Privatsphäre besser zu schützen?*

Sensibilität.

- *[Für jedes sensible Verhalten separat fragen:] Aus Ihrer Perspektive, wie unwohl würde sich ein Wissenschaftler/eine Wissenschaftlerin fühlen, wenn sie dieses Verhalten zugeben müssten?*
- *[Für jedes sensible Verhalten separat fragen:] Wie würden Sie dieses Verhalten ethisch beurteilen? Würde ein Wissenschaftler/eine Wissenschaftlerin sich falsch/unethisch verhalten, wenn er/sie dies tun würden?*
- *[Für alle Verhalten:] Könnten Sie bitte die sechs Verhaltensweisen nach ihrer Sensibilität ordnen?*

Sonstige.

- *Haben Sie Verbesserungsvorschläge?*

Verabschiedung

Ich bedanke mich recht herzlich, dass Sie sich die Zeit genommen haben, um dieses Gespräch mit mir zu führen. Ich möchte Ihnen nun den Umschlag mit Ihrem Fragebogen zurückgeben. Sie dürfen damit nun tun, was Sie möchten. Wie bereits angedeutet, bin ich an Ihren tatsächlichen Antworten nicht interessiert. Wenn Sie den Umschlag nicht behalten möchten, schlage ich vor, dass wir diesen nun gemeinsam vernichten.

Appendix C

Questionnaire (English)

Introduction

Good morning/afternoon/evening my name is ...from Thanks for agreeing to participate in this study.

I would like to begin by explaining the procedure to you in more detail. I will give you a paper questionnaire, first, and I would like to ask you to fill this in. The questionnaire includes relatively novel question techniques that aim at better preserving your privacy. Please read the instruction carefully and then answer the questions. Once you have completed the questionnaire, I would like you to put the questionnaire in this envelope and to seal it securely.

Next, I will ask you a couple of questions about your understanding of the questions. I am particularly interested to find out how you came up with your answers and if you had any concerns when answering any of the questions. For the success of our study, it is important that you answer truthfully.

You will find all information about this study on the participant information form [Hand over Participant Information Sheet]. As this is a scientific research project, I need to obtain written consent that you agree to take part in this study [Hand over Participant Consent Form]. If you are happy to participate in this study, could you please sign this form for me. This information will be filed with the ethics commission only.

Of course, this survey and our conversation afterwards are completely confidential. In order to capture all information accurately, I would like to record our conversation. If you are happy for me to do this, could you please confirm this by signing at the bottom of the consent form again.

Thank you very much. Let's begin.

Q1.

For this question, you are asked two questions in one block. Please start thinking about how you would answer each question individually (either yes or no), but do not write this down. Depending on how you would answer these two questions, please indicate whether the answer is (A) or (B) following the instructions below:

If your answer to both question is no or the answer to both questions is yes, please indicate this by selecting answer (A). If your answer to one of the questions is yes and it is no to the other, please indicate this by selecting answer (B).

Your privacy remains protected, as I do not know your answer to the individual questions.

- Is your mother's birthday in January or February?
- Have you ever claimed first authorship for a paper knowing that someone else contributed more than you?

What is the answer to both questions?

- ☐ Both questions yes or both questions no
- ☐ One question yes and the other one no

Q2.

Once again, you are asked two questions in one block. Please start thinking about how you would answer each question individually (either yes or no), but do not write this down. Depending on how you would answer these two questions, please indicate whether the answer is (A) or (B) following the instructions below:

If your answer to both question is no or the answer to both questions is yes, please indicate this by selecting answer (A). If your answer to one of the questions is yes and it is no to the other, please indicate this by selecting answer (B).

Your privacy remains protected, as I do not know your answer to the individual questions.

- Is your father's birthday in January or February?
- Have you ever intentionally used someone else's ideas without giving them credit?

What is the answer to both questions?

- ☐ Both questions yes or both questions no
- ☐ One question yes and the other one no

Q3.

I will read out a list of activities that some researchers do, but other do *not* do. Please tell me, *how many* of these things you have *ever* done before. Once again please don't tell me which ones you have done, but only how many of them you have *ever* done:

- On a typical work day, I commute more than 50km one way.
- I regularly participate in my institution's research meetings.
- I have strategically excluded analyses from my paper to increase the success of a publication.
- I have a personal subscription for at least two scientific journal.
- I predominantly use Windows for work.

Please write down below *how many* of the activities you have ever done.

Q4.

Once again, I will read out a list of activities that some researchers do, but other do *not* do. Please tell me, *how many* of these things you have *ever* done before. Once again please don't tell me which ones you have done, but only how many of them you have *ever* done:

- I have lunch with colleagues every week.
- In the past 12 month, I have presented my work in more than 3 languages.
- I have intentionally manipulated empirical data to support my research question.
- I regularly take handwritten notes for scientific purposes.
- I have a personal subscription for the print version of at least two newspapers.

Please write down below *how many* of the activities you have ever done.

Q5.

For this question, you are asked two questions in one block again. Please start thinking about how you would answer each question individually (either yes or no), but do not write this down. Depending on how you would answer these two questions, please indicate whether the answer is (A) or (B) this time following the new instructions below:

If your answer to both question is no, please indicate this by selecting answer (A). If your answer to at least one of the questions is yes, please indicate this by selecting answer (B).

Your privacy remains protected, as I do not know your answer to the individual questions.

- Is your father's birthday the 1st, 2nd, 3rd, 4th or 5th of the month?
- Have you ever changed a hypothesis when you already knew the empirical results to give the impression that it was supported?

What is the answer to both questions?

- ☐ both with no
- ☐ at least one with yes

Q6.

For this question, you are asked two questions in one block again. Please start thinking about how you would answer each question individually (either yes or no), but do not write this down. Depending on how you would answer these two questions, please indicate whether the answer is (A) or (B) this time following the new instructions below:

If your answer to both question is no, please indicate this by selecting answer (A). If your answer to at least one of the questions is yes, please indicate this by selecting answer (B).

Your privacy remains protected, as I do not know your answer to the individual questions.

- Is your mother's birthday the 1st, 2nd, 3rd, 4th or 5th of the month?
- Have you ever withheld results because they contradicted your previous research?

What is the answer to both questions?

- ☐ both with no
- ☐ at least one with yes

Could you please put the paper questionnaire in this envelope [Hand over envelope] and seal it securely. I will then ask you a couple of questions.

Probes

This was an unusual way of asking survey questions. I will now show you these questions again [Display showcard with questions]. Just for emphasis, I am interested in what you were thinking when you heard/read these questions. Please just think out loud when answering the next couple of questions.

Understanding.

- *What do you think, why we have asked the question in this way?*
- *Please look at the three different question formats again. What did you think, when coming up with your answers? How did you arrive at your answers?*
- *Did you have any problems with the instructions? [Do you have any suggestions for improvements?]*
- *Could you repeat the instructions but in your own words?*
- *Could you explain to me why your answers are better protected with these methods of asking the questions?*
- *What do you think, how could I use your answers in a meaningful way for data analysis?*
- *[For each method separately:] Do you think there is an answer that would allow nobody to assume that you did [Insert sensitive item here]?*

Honesty.

- *Do you think other researchers would answer honestly?*
- *Did you answer truthfully?*
- *Which method do you think protects your privacy best?*
- *What about the other methods? Is there a method you feel uncomfortable with?*
- *Did you miss anything in the instructions that would help you to know that your privacy is protected?*

Sensitivity.

- *[For each sensitive behavior:] From your point of view, how uneasy would a researcher feel admitting this behavior?*
- *[For each sensitive behavior:] How would you evaluate the behavior from an ethical point of view? Would a researcher be wrong/very unethical if they did this?*
- *[Looking at all behaviors:] Could you please rank the six behaviors according to how sensitive they are?*

Other.

- *Do you have any recommendations for improvement?*

Goodbye

I really appreciate that you took the time to participate in this interview. I would like to hand you back your envelope now. You can do whatever you want to do with it now. As I said earlier, I am not at all interested in the answers you gave. If you do not want to keep it, I suggest we destroy the envelope now.

Appendix D Advance Letter



SENDER ADDRESS

DEPARTMENT

Telephone:
Email:
Webpage:
Building:

RECEIVER ADDRESS

London, DATE

Dear PROSPECTIVE PARTICIPANT

Invitation to interview about special survey questions for sensitive topics

I would like to invite you to participate in a short lunchtime interview (10 to 15 minutes) on the functioning of special survey question formats in the field of Higher Education research.

This interview is part of a collaborative research project that I conduct with CO-PIs (ORGANISATION CO-PIs). We want to find out how academics at different levels of their career understand these question formats. I am in charge of the data collection at ORGANISATION (Ethics Reference: Soc-REC/80025566/21-03-17).

You have been selected on the basis of your academic field, career status, and gender. We selected you by browsing the publicly available information on your department's homepage. Your participation in this research project is, of course, completely voluntary. All answers you provide will be anonymous and treated fully confidentially.

I will try to reach you by telephone in the next few days in order to give you the opportunity to ask me any questions you may have regarding this project. Ideally, I would also like to arrange a convenient time for us to meet and to conduct the interview. Do not hesitate to contact me (TELEPHONE NUMBER; EMAIL ADDRESS) in case you have any further questions in the meantime.

Yours faithfully,

SIGNATURE

Appendix E

Participant Information Sheet



The Applicability of Randomised Response Techniques (RRT) and the Item Count Technique (ICT) in the Field of Higher Education and Science Studies

Participant Information Form

We would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

What is the purpose of the study?

This is a collaborative research project [redacted] investigating participants' understanding of special survey question formats for sensitive questions.

Why have I been invited?

We want to understand how academics at all levels understand these different survey question formats. We have randomly selected a small sample of participants according to set quotas regarding the academic field, researchers' status, and their gender.

Do I have to take part?

Your participation in this research project is completely voluntary. You can withdraw from part of the research or the project as a whole at any stage without being penalised or disadvantaged in any way. It is entirely up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a consent form. This is a standard ethical procedure. If you decide to take part, you are still free to withdraw at any time and without giving a reason. The data will be fully anonymised.

What will happen if I take part?

- You will take part in so-called cognitive interview, in which we would like to find out more about how you understand and interpret a small number of survey questions.
- This should not take up more than 15-20 minutes of your time.
- Your participation in the study will end with the end of this interview.
- We will not collect any personal information.
- The research will take place at [redacted]

What do I have to do?

- You are asked to voice your reasoning about a series of survey questions.
- This includes to tell us how you understand a particular question format, whether you feel that it better protects your privacy and anonymity, as well as if you find particular questions particularly sensitive to ask in the Higher Education environment.

What are the possible disadvantages and risks of taking part?

- There are no risks or disadvantages associated with this research.
- We ensure full anonymity and confidentiality of your answers.
- This research is for scientific purposes only.

What are the possible benefits of taking part?

This research is associated with indirect benefits. For instance, this research will inform future large n survey data collections and lead to a better protection of survey respondents' anonymity and privacy in surveys.

What will happen when the research study stops?

Once the research study has finished, the anonymized data will continue to be stored on a secure server at [REDACTED] for a max. of 10 years. This is in accordance with the ethical guidelines at [REDACTED]. The data will be deleted and the device physically destroyed after this deadline.

Will my taking part in the study be kept confidential?

- Only the local investigator, [REDACTED], will have access to your raw data.
- An anonymized data file will be made available via secure connections to the other investigators, but will not be shared with any third parties.
- If you have given consent to audio recordings, these will be stored in a password protected folder on a secure server. Only your local investigator will have access to these data.
- The data will not be achieved or shared with third parties.
- This project collects data for research purposes only, none of the information shared will result in any risk of disclosure or other consequences for the participant.

What will happen to results of the research study?

The results of this study will inform a conference presentation and research publication. This research will not mention any details that may allow a third party to re-identify participants. If you are interested in the results of this research, you can reach your local investigator by email:

[REDACTED] or telephone: [REDACTED]

What will happen if I do not want to carry on with the study?

You can withdraw from this study at any point in time. Withdrawal will not result in any penalty or consequences for you.

What if there is a problem?

If you have any further enquiries about this research or encounter any problems, you can reach your local investigator by email: [REDACTED] or telephone: [REDACTED]

If you have any problems, concerns or questions about this study, you should ask to speak to a member of the research team. If you remain unhappy and wish to complain formally, you can do this through City's complaints procedure. To complain about the study, you need to phone [REDACTED]. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the reference of the project is: [REDACTED]

You could also write to the Secretary at:

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

City holds insurance policies which apply to this study. If you feel you have been harmed or injured by taking part in this study you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone's negligence, then you may have grounds for legal action. This study has been approved by City Social Science Research Ethics Committee.

Appendix F

Consent to Participate Form



The Applicability of Randomised Response Techniques (RRT) and the Item Count Technique (ICT) in the Field of Higher Education and Science Studies

Participant Consent Form

I confirm that the aim of the project has been explained to me and that I have understood that this research will involve being cognitively interviewed by a researcher.

I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be published. The identifiable data will not be shared with any other organisation. Only fully anonymous transcripts of the data may be passed on to the core scientific team for analysis.

My participation in this research is completely voluntary. I understand that I can withdraw from part of the project or the project as a whole at any stage of the project without being penalised or disadvantaged in any way.

I agree to City recording and processing the information provided on this form about me. I understand that this information will be used only for the purposes set out in this statement and my consent is conditional on City complying with its duties and obligations under the Data Protection Act 1998.

I agree to take part in the above study.

Name of Participant

Signature

Date

I allow the interview to be audiotaped.

Name of Participant

Signature

Date

Name of Researcher

Signature

Date